Consistency and Robustness Properties of SVMs for Heavy-Tailed Distributions

Arnout Van Messem Andreas Christmann

Vrije Universiteit Brussel
UNIVERSITÄT BAYREUTH

17th Annual Meeting of the Belgian Statistical Society,
October 14-16, 2009, Lommel, Belgium

Introduction

Known

Support Vector Machines (SVMs) are **consistent** and **robust**, if based on Lipschitz continuous loss and bounded kernel.

Christmann & Van Messem '08
Steinwart & Christmann '08
Christmann & Steinwart '07

Question

Can the assumptions $f \in L_1(\mathrm{P}_X)$ and $\int |Y| \, d\mathrm{P} < \infty$ be weakened?

(both for regression and classification problems)

Support Vector Machines

Definition

$$f_{L,\mathrm{P},\lambda} := \arg \inf_{f \in \mathcal{H}} \mathbb{E}_{\mathrm{P}} L(X, Y, f(X)) + \lambda \|f\|_{\mathcal{H}}^2$$

- $\mathcal{X} \subseteq \mathbb{R}^d$ closed, $\mathcal{Y} \subseteq \mathbb{R}$ closed, $\mathcal{X} \neq \varnothing$, $\mathcal{Y} \neq \varnothing$
- (X_i, Y_i) i.i.d. $\sim \mathrm{P} \in \mathcal{M}_1(\mathcal{X} \times \mathcal{Y})$, P (totally) unknown
- $Y_i | x_i$ depends on an *unknown* function $f : \mathcal{X} \to \mathbb{R}$
- **RKHS** $\mathcal{H} \rightleftarrows$ **kernel** $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, k **measurable**
- **Loss function** $L : \mathcal{X} \times \mathcal{Y} \times \mathbb{R} \to [0, \infty)$, $L(x, y, f(x))$
- $\lambda > 0$ regularization parameter
- $f_{L,\mathrm{D},\lambda}$, where D is empirical distribution for data set $D = ((x_1, y_1), \ldots, (x_n, y_n)) \in (\mathcal{X} \times \mathcal{Y})^n$

Support Vector Machines

Notions

- L is called **convex**, **continuous**, **Lipschitz continuous**, **differentiable**, if L has this property w.r.t. 3^{rd} argument
- k is called **bounded**, if $||k||_\infty := \sqrt{\sup_{x\in\mathcal{X}} k(x,x)} < \infty$
- $\Phi : \mathcal{X} \to \mathcal{H}$, $\Phi(x) := k(\cdot, x)$, is called **canonical feature map**
- Reproducing property:

$$f(x) = \langle f, k(x,\cdot)\rangle_{\mathcal{H}} \quad \forall\, f \in \mathcal{H}, \forall\, x \in \mathcal{X}.$$

Risk

Definitions

Risk	$\mathcal{R}_{L,\mathrm{P}}(f)$	$\mathbb{E}_{\mathrm{P}} L(X, Y, f(X))$
Bayes risk	$\mathcal{R}^*_{L,\mathrm{P}}$	$\inf_{f:\mathcal{X}\to\mathbb{R} \text{ measurable}} \mathcal{R}_{L,\mathrm{P}}(f)$
Bayes function	$f^*_{L,\mathrm{P}}$	$\arg\inf_{f:\mathcal{X}\to\mathbb{R} \text{ measurable}} \mathcal{R}_{L,\mathrm{P}}(f)$

Questions

Under which conditions on $\mathcal{X}$, $\mathcal{Y}$, L, $\mathcal{H}$, and k do we have:

1. $f_{L,\mathrm{P},\lambda}$: existence, uniqueness, representation
2. Universal consistency to Bayes risk/function, i.e., $\forall\ \mathrm{P}$

$$\mathcal{R}_{L,\mathrm{P}}(f_{L,\mathrm{D},\lambda}) \xrightarrow{\mathrm{P}} \mathcal{R}^*_{L,\mathrm{P}} \text{ for } |D| = n \to \infty$$
$$f_{L,\mathrm{D},\lambda} \xrightarrow{\mathrm{P}} f^*_{L,\mathrm{P}} \text{ for } |D| = n \to \infty$$

3. Robustness of $f_{L,\mathrm{P},\lambda}$?

Shifted loss function

Loss function $L : \mathcal{X} \times \mathcal{Y} \times \mathbb{R} \to [0, \infty)$ measurable

Definition

$L^\star : \mathcal{X} \times \mathcal{Y} \times \mathbb{R} \to \mathbb{R}$ with

$$L^\star(x, y, t) := L(x, y, t) - L(x, y, 0).$$

Huber, 1967

$L^\star$ can be negative!

Properties

- L (strictly) convex, then $L^\star$ (strictly) convex.
- L Lipschitz continuous, then $L^\star$ Lipschitz continuous.

Shifted loss function

Conditions for finite risk

For L Lipschitz continuous

- $\mathcal{R}_{L,\mathrm{P}}(f) < \infty$ if $f \in L_1(\mathrm{P}_X)$ and $\mathbb{E}_{\mathrm{P}}|Y| < \infty$.
- $\mathcal{R}_{L^\star,\mathrm{P}}(f) < \infty$ if $f \in L_1(\mathrm{P}_X)$.

Equality of SVMs

If $f_{L,\mathrm{P},\lambda}$ exists, then $f_{L^\star,\mathrm{P},\lambda} = f_{L,\mathrm{P},\lambda}$.

Properties

- If L Lipschitz continuous, then

$$|\mathcal{R}_{L^\star,\mathrm{P}}(f)| \leq |L|_1 \mathbb{E}_{\mathrm{P}_X}|f(X)|.$$

$$|\mathcal{R}^{reg}_{L^\star,\mathrm{P},\lambda}(f)| \leq |L|_1 \mathbb{E}_{\mathrm{P}_X}|f(X)| + \lambda \left\|f\right\|^2_{\mathcal{H}}.$$

- If L Lipschitz continuous and $f_{L^\star,\mathrm{P},\lambda}$ exists, then

$$\|f_{L^\star,\mathrm{P},\lambda}\|^2_{\mathcal{H}} \leq \lambda^{-1} \min\left\{|L|_1 \mathbb{E}_{\mathrm{P}_X}|f_{L^\star,\mathrm{P},\lambda}(X)|, \mathcal{R}_{L,\mathrm{P}}(0)\right\}.$$

  If additionally k is bounded, then $\|f_{L^\star,\mathrm{P},\lambda}\|_{\mathcal{H}} < \infty$.

Existence and Uniqueness of SVM solution

Uniqueness

- L convex and $\mathcal{R}_{L^\star,\mathrm{P}}(f) < \infty$ for some $f \in \mathcal{H}$ and $\mathcal{R}_{L^\star,\mathrm{P}}(f) > -\infty$ for all $f \in \mathcal{H}$
  OR
- L is convex, Lipschitz continuous and $f \in L_1(\mathrm{P}_X)$.

Then, for all $\lambda > 0$, there **exists at most one** SVM $f_{L^\star,\mathrm{P},\lambda}$.

Existence

- L convex, Lipschitz continuous,
- $\mathcal{H}$ RKHS of a bounded measurable kernel k.

Then, for all $\mathrm{P} \in \mathcal{M}_1(\mathcal{X} \times \mathcal{Y})$ and for all $\lambda > 0$, there **exists** an SVM solution $f_{L^\star,\mathrm{P},\lambda}$.

Representation

Theorem

- L convex, Lipschitz continuous loss function,
- k bounded, measurable kernel with separable RKHS $\mathcal{H}$.

Then, for all $\mathrm{P} \in \mathcal{M}_1(\mathcal{X} \times \mathcal{Y})$ and for all $\lambda > 0$, there exists an $h \in \mathcal{L}_\infty(\mathrm{P})$ with

$$\begin{aligned} h(x,y) &\in \partial L^\star(x,y,f_{L^\star,\mathrm{P},\lambda}(x))\,,\ \forall\,(x,y) \\ \|h\|_\infty &\leq |L^\star|_1 = |L|_1 \\ f_{L^\star,\mathrm{P},\lambda} &= -\frac{1}{2\lambda}\mathbb{E}_\mathrm{P}(h\Phi) \\ \|f_{L^\star,\mathrm{P},\lambda} - f_{L^\star,\mathrm{Q},\lambda}\|_\mathcal{H} &\leq \frac{1}{\lambda}\,\|\mathbb{E}_\mathrm{P}(h\Phi) - \mathbb{E}_\mathrm{Q}(h\Phi)\|_\mathcal{H}\,,\ \forall\,\mathrm{Q}. \end{aligned}$$

Consistency

Theorem

- L convex, Lipschitz continuous loss function,
- $\mathcal{H}$ separable RKHS of a bounded, measurable kernel k,
- $\mathcal{H}$ dense in $L_1(\mu)$ for all distributions μ on $\mathcal{X}$,
- (λ_n) sequence of strictly positive numbers with $\lambda_n \to 0$.

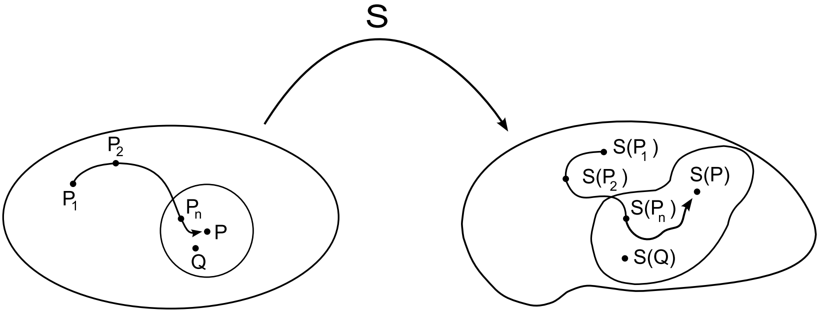
Then, for all $\mathrm{P} \in \mathcal{M}_1(\mathcal{X} \times \mathcal{Y})$ and all D with $|D| = n$,

1. if $\lambda_n^2 n \to \infty$, then $\mathcal{R}_{L^\star,\mathrm{P}}(f_{L^\star,\mathrm{D},\lambda_n}) \xrightarrow{\mathrm{P}} \mathcal{R}^*_{L^\star,\mathrm{P}}$.
2. if $\lambda_n^{2+\delta} n \to \infty$ for some $\delta \in (0,\infty)$, then
$$\mathcal{R}_{L^\star,\mathrm{P}}(f_{L^\star,\mathrm{D},\lambda_n}) \xrightarrow{\mathrm{a.s.}} \mathcal{R}^*_{L^\star,\mathrm{P}} .$$
3. if $L = L_\tau$ pinball loss: $d(f_{L^\star,\mathrm{D},\lambda_n}, f^*_{L_\tau,\mathrm{P}}) \to 0.$

   d is a metric describing convergence in probability.

Robustness

1. What if (X_i, Y_i) i.i.d. $\sim \mathrm{P}$, $\mathrm{P} \in \mathcal{M}_1$ unknown is invalid?
2. What is the impact on $S : \mathrm{P} \mapsto f_{L^*,\mathrm{P},\lambda}$?



Derivatives and Influence Functions

$$\begin{tikzcd}
 & \text{Hadamard} \arrow[r, Rightarrow] & \text{Gâteaux} \arrow[dr, Rightarrow] & \\
\text{Fréchet} \arrow[ur, Rightarrow] \arrow[dr, Rightarrow] & & & \mathrm{IF} \\
 & \text{Bouligand} \arrow[r, Rightarrow] & \mathrm{BIF} \arrow[ur, Rightarrow] &
\end{tikzcd}$$

Christmann & Van Messem (2008)

Notation: ∇^F, ∇^G, ∇^B, ∇_3^B, etc.

Property: $\nabla_3^F L^\star = \nabla_3^F L$, $\nabla_3^B L^\star = \nabla_3^B L$

Influence Function

Definition (Hampel, '68, Hampel et al. '86)

The **influence function** (IF) of a function $S : \mathcal{M}_1 \to \mathcal{H}$ for a distribution P is given by

$$\mathrm{IF}(z; S, \mathrm{P}) := \lim_{\varepsilon \downarrow 0} \frac{S\big((1-\varepsilon)\mathrm{P} + \varepsilon \delta_z\big) - S(\mathrm{P})}{\varepsilon},$$

in those $z := (x, y) \in \mathcal{X} \times \mathcal{Y}$ where this limit exists.

If $\nabla^G(z; S, \mathrm{P})$ exists: $\nabla^G = \mathrm{IF}$ and IF is linear and continuous

Goal: Bounded IF
Problem: Loss function L often not Fréchet-differentiable

Bouliganud Influence Function

Definition (C&VM '08)

The **Bouligand influence function** (BIF) of a function $S : \mathcal{M}_1 \to \mathcal{H}$ for a distribution P in the direction of a distribution $\mathrm{Q} \neq \mathrm{P}$ is the special Bouligand-derivative

$$\lim_{\varepsilon \downarrow 0} \frac{\left\| S\big((1-\varepsilon)\mathrm{P} + \varepsilon \mathrm{Q}\big) - S(\mathrm{P}) - \mathrm{BIF}(\mathrm{Q}; S, \mathrm{P}) \right\|_{\mathcal{H}}}{\varepsilon} = 0$$

(if it exists).

If BIF exists and $\mathrm{Q} = \delta_z$: IF exists and $\mathrm{BIF} = \mathrm{IF}$

Goal: Bounded BIF

Result for IF

Assumptions

- $\mathcal{H}$ is RKHS with **bounded**, continuous kernel k
- L **convex** and **Lipschitz continuous**
- $\nabla_3^F L(x, y, \cdot)$ and $\nabla_{3,3}^F L(x, y, \cdot)$ continuous with
  $\kappa_1 := \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \left\|\nabla_3^F L(x, y, \cdot)\right\|_\infty \in (0, \infty)$,
  $\kappa_2 := \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \left\|\nabla_{3,3}^F L(x, y, \cdot)\right\|_\infty < \infty$

Theorem IF

Then $\mathrm{IF}(z; S, \mathrm{P})$ with $S(\mathrm{P}) := f_{L^\star,\mathrm{P},\lambda}$ and $z := (x, y)$

1. exists,
2. equals

$$\mathbb{E}_{\mathrm{P}} \nabla_3^F L^\star(X, Y, f_{L^\star,\mathrm{P},\lambda}(X)) T^{-1} \Phi(X)$$
$$- \nabla_3^F L^\star(x, y, f_{L^\star,\mathrm{P},\lambda}(x)) T^{-1} \Phi(x) \,,$$

where $T : \mathcal{H} \to \mathcal{H}$ with $T(\cdot) :=$
$2\lambda \, \mathrm{id}_{\mathcal{H}}(\cdot) + \mathbb{E}_{\mathrm{P}} \nabla_{3,3}^F L^\star(X, Y, f_{L^\star,\mathrm{P},\lambda}(X)) \langle \Phi(X), \cdot \rangle_{\mathcal{H}} \Phi(X)$,

3. is bounded.

Bounds for bias

Maxbias and IF

- $\mathcal{H}$ is separable RKHS with **bounded**, measurable kernel k
- L **convex** and **Lipschitz continuous**

Then, for all $\lambda > 0$, all $\varepsilon \in [0,1]$ and all $\mathrm{P}, \mathrm{Q} \in \mathcal{M}_1(\mathcal{X} \times \mathcal{Y})$

$$\left\| f_{L^\star,(1-\varepsilon)\mathrm{P}+\varepsilon\mathrm{Q}} - f_{L^\star,\mathrm{P},\lambda} \right\|_{\mathcal{H}} \leq c_{\mathrm{P},\mathrm{Q}}\varepsilon,$$

where $c_{\mathrm{P},\mathrm{Q}} = \lambda^{-1}\|k\|_\infty |L|_1 \, \|\mathrm{P}-\mathrm{Q}\|_{\mathcal{M}}$.

- $\mathrm{Q} = \delta_z$ with $z := (x,y)$
- $\mathrm{IF}(z;S,\mathrm{P})$ with $S(\mathrm{P}) := f_{L^\star,\mathrm{P},\lambda}$ exists

Then $$\|\mathrm{IF}(z;S,\mathrm{P})\|_{\mathcal{H}} \leq c_{\mathrm{P},\delta_z}.$$

Result for BIF

Assumptions

- $\mathcal{H}$ is RKHS with **bounded**, continuous kernel k
- L **convex** and **Lipschitz continuous** with $|L|_1 \in (0, \infty)$
- $\nabla_3^B L(x, y, \cdot)$ and $\nabla_{3,3}^B L(x, y, \cdot)$ measurable with $\kappa_1 := \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \left\| \nabla_3^B L(x, y, \cdot) \right\|_\infty \in (0, \infty)$, $\kappa_2 := \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \left\| \nabla_{3,3}^B L(x, y, \cdot) \right\|_\infty < \infty$

Assumptions

- $\delta_1 > 0$, $\delta_2 > 0$
- $\mathcal{N}_{\delta_1}(f_{L^\star,\mathrm{P},\lambda}) := \{f \in \mathcal{H} : \|f - f_{L^\star,\mathrm{P},\lambda}\|_{\mathcal{H}} < \delta_1\}$
- $\lambda > \frac{1}{2}\kappa_2 \|k\|_\infty^{\,3}$ $\quad$ ($\kappa_2 = 0$ for eps-insensitive and pinball)
- $\mathrm{P} \neq \mathrm{Q}$, probability measures on $\mathcal{X} \times \mathcal{Y}$
- Define $G : (-\delta_2, \delta_2) \times \mathcal{N}_{\delta_1}(f_{L^\star,\mathrm{P},\lambda}) \to \mathcal{H}$,

$$G(\varepsilon, f) := 2\lambda f + \mathbb{E}_{(1-\varepsilon)\mathrm{P}+\varepsilon\mathrm{Q}} \nabla_3^B L^\star(X, Y, f(X))\Phi(X)$$

- $G(0, f_{L^\star,\mathrm{P},\lambda}) = 0$ and $\nabla_2^B G(0, f_{L^\star,\mathrm{P},\lambda})$ is **strong**

Theorem BIF

Then $\mathrm{BIF}(\mathrm{Q}; S, \mathrm{P})$ with $S(\mathrm{P}) := f_{L^\star,\mathrm{P},\lambda}$ and $\mathrm{Q} \neq \mathrm{P} \in \mathcal{M}_1$

1. exists,
2. equals

$$T^{-1}\Big(\mathbb{E}_{\mathrm{P}} \nabla_3^B L^\star(X, Y, f_{L^\star,\mathrm{P},\lambda}(X))\Phi(X) \\ -\mathbb{E}_{\mathrm{Q}} \nabla_3^B L^\star(X, Y, f_{L^\star,\mathrm{P},\lambda}(X))\Phi(X)\Big),$$

   where $T : \mathcal{H} \to \mathcal{H}$ with $T(\cdot) :=$
   $2\lambda \,\mathrm{id}_{\mathcal{H}}(\cdot) + \mathbb{E}_{\mathrm{P}} \nabla_{3,3}^B L^\star(X, Y, f_{L^\star,\mathrm{P},\lambda}(X)) \langle \Phi(X), \cdot \rangle_{\mathcal{H}} \Phi(X)$,
3. is bounded.

Conclusions

SVMs based on $L^\star(x, y, t) := L(x, y, t) - L(x, y, 0)$

1. Weaker assumption on P: only $f \in L_1(\mathrm{P}_X)$ is needed
   e.g. f bounded and $\mathcal{X} \subset \mathbb{R}^d$ bounded
2. Existence and uniqueness of $f_{L^\star,\mathrm{P},\lambda}$
3. Representation of SVM solution
4. Consistency of risk and SVM solution
5. Robustness
   - Existence of IF and BIF
   - $\mathrm{IF}(\mathrm{Q}; S, \mathrm{P})$ bounded if $\nabla_3^F L$, $\nabla_{3,3}^F L$ and k continuous and bounded
   - $\mathrm{BIF}(\mathrm{Q}; S, \mathrm{P})$ bounded if $\nabla_3^B L$, $\nabla_{3,3}^B L$ measurable and bounded as well as k continuous and bounded
   - Bounds for bias

References

- Christmann, Van Messem & Steinwart (2009). *Statistics and Its Interface*, **2**, 311-327.
- Christmann & Van Messem (2008). *Journal of Machine Learning Research*, **9**, 915-936.
- Steinwart & Christmann (2008). *Support Vector Machines.* Springer, New York.
- Christmann & Steinwart (2007). *Bernoulli*, **13**, 799-819.
- Hampel (1974). *J. Amer. Statist. Assoc.*, **69**, 383-393.
- Huber (1967). *Proceedings of the 5^{th} Berkeley Symposium.*
- Koenker (2005). *Quantile regression.* Cambridge University Press.
- Schölkopf & Smola (2002). *Learning with kernels.* MIT Press.
- Vapnik (1998). *Statistical learning theory.* Wiley.

Reason

Conditions for finite risk

For L Lipschitz continuous

- $\mathbb{E}_{\mathrm{P}} L(X, Y, f(X)) < \infty$ if $f \in L_1(\mathrm{P}_X)$ and $Y \in L_1(\mathrm{P}_{Y|x})$.

$$\mathcal{R}_{L,\mathrm{P}}(f) \leq |L|_1 \left(\int_{\mathcal{X}} |f(x)| d\mathrm{P}_X(x) + \int_{\mathcal{X}} \int_{\mathcal{Y}} |y| d\mathrm{P}(y|x)\, d\mathrm{P}_X(x) \right)$$

- $\mathbb{E}_{\mathrm{P}} L^{\star}(X, Y, f(X)) < \infty$ if $f \in L_1(\mathrm{P}_X)$.

$$\mathcal{R}_{L,\mathrm{P}}(f) \leq |L|_1 \int_{\mathcal{X}} |f(x)|\, d\mathrm{P}_X(x)$$

Sketch of proof for IF

- $G(\varepsilon, f) := 2\lambda f + \mathbb{E}_{(1-\varepsilon)\mathrm{P}+\varepsilon\Delta_z} \nabla_2^F L^*(Y, f(X))\Phi(X)$
- $G(\varepsilon, f) = \nabla_2^F \mathcal{R}^{reg}_{L^*,(1-\varepsilon)\mathrm{P}+\varepsilon\Delta_z,\lambda}(f), \qquad \varepsilon \in [0,1]$
- $G(\varepsilon, f)$ fulfills conditions of a standard implicit function theorem on Banach spaces

Sketch: Proof for IF

For the proof of the theorem about the IF we showed:

i. $G(0,f)=0 \Leftrightarrow f=f_{L^*,\mathrm{P},\lambda}$.

ii. G continuously F-differentiable.

iii. $\frac{\partial G}{\partial \mathcal{H}}(0,f_{L^*,\mathrm{P},\lambda})$ invertible.

iv. Then there exist $\delta>0$, a neighborhood $\mathcal{N}_\delta(f_{L^*,\mathrm{P},\lambda}) := \{f\in\mathcal{H};\, \|f-f_{L^*,\mathrm{P},\lambda}\|_{\mathcal{H}}<\delta\}$, and a function $f^*:(-\delta,\delta)\to\mathcal{N}_\delta(f_{L^*,\mathrm{P},\lambda})$ satisfying

iv.1) $f^*(0)=f_{L^*,\mathrm{P},\lambda}$.

iv.2) It holds

$$\nabla^F f^*(0) = -\left(\nabla_2^F G(0,f_{L^*,\mathrm{P},\lambda})\right)^{-1} - \nabla_1^B G(0,f_{L^*,\mathrm{P},\lambda}).$$

Sketch of proof for BIF

- $\nabla_2^B L^\star(Y, f(X)) = \nabla_2^B L(Y, f(X))$ hence

$$G(\varepsilon, f) = 2\lambda f + \mathbb{E}_{(1-\varepsilon)\mathrm{P}+\varepsilon\mathrm{Q}} \nabla_2^B L(Y, f(X))\Phi(X)$$

- $G(\varepsilon, f) = \nabla_2^B \mathcal{R}^{reg}_{L^\star,(1-\varepsilon)\mathrm{P}+\varepsilon\mathrm{Q},\lambda}(f), \qquad \varepsilon \in [0, 1]$
- $G(\varepsilon, f)$ fulfills the conditions of Robinson's (1991) implicit function theorem on Bouligand-derivatives for non-smooth functions in Banach or normed linear spaces

$\Rightarrow$ Rest of proof uses same arguments as Christmann & Van Messem (2008).

Sketch: Proof for BIF

For the proof of the theorem about the BIF we showed:

i. For some χ and each $f \in \mathcal{N}_{\delta_1}(f_{L^*,\mathrm{P},\lambda})$, $G(\cdot\,, f)$ is Lipschitz continuous on $(-\delta_2, \delta_2)$ with Lipschitz constant χ.

ii. G has partial B-derivatives with respect to ε and f at $(0, f_{L^*,\mathrm{P},\lambda})$.

iii. $\nabla_2^B G(0, f_{L^*,\mathrm{P},\lambda})\big(\mathcal{N}_{\delta_1}(f_{L^*,\mathrm{P},\lambda}) - f_{L^*,\mathrm{P},\lambda}\big)$ is a neighborhood of $0 \in \mathcal{H}$.

iv. $\delta\big(\nabla_2^B G(0, f_{L^*,\mathrm{P},\lambda}), \mathcal{N}_{\delta_1}(f_{L^*,\mathrm{P},\lambda}) - f_{L^*,\mathrm{P},\lambda}\big) \;=:\; d_0 > 0$.

v. For each $\xi > d_0^{-1}\chi$ there exist $\delta_3, \delta_4 > 0$, a neighborhood $\mathcal{N}_{\delta_3}(f_{L^*,\mathrm{P},\lambda}) := \{f \in \mathcal{H}; \|f - f_{L^*,\mathrm{P},\lambda}\|_{\mathcal{H}} < \delta_3\}$, and a function $f^* : (-\delta_4, \delta_4) \to \mathcal{N}_{\delta_3}(f_{L^*,\mathrm{P},\lambda})$ satisfying

- **v.1)** $f^*(0) = f_{L^*,\mathrm{P},\lambda}$.
- **v.2)** $f^*(\cdot)$ is Lipschitz continuous on $(-\delta_4, \delta_4)$ with Lipschitz constant $|f^*|_1 = \xi$.
- **v.3)** For each $\varepsilon \in (-\delta_4, \delta_4)$ is $f^*(\varepsilon)$ the unique solution of $G(\varepsilon, f) = 0$ in $(-\delta_4, \delta_4)$.
- **v.4)** It holds $\nabla^B f^*(0)(u) = \left(\nabla_2^B G(0, f_{L^*,\mathrm{P},\lambda})\right)^{-1} \left(-\nabla_1^B G(0, f_{L^*,\mathrm{P},\lambda})(u)\right)$.